ILIAD: A COMPUTER PROGRAM FOR MANAGING PERSONAL BIBLIOGRAPHIC KNOWLEDGE BASES

by

James Carr Smith

A thesis submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Master of Science

Department of Medical Informatics

The University of Utah

June 1986

THE UNIVERSITY OF UTAH GRADUATE SCHOOL

SUPERVISORY COMMITTEE APPROVAL

of a thesis submitted by

James Carr Smith

This thesis has been read by each member of the following supervisory committee and by majority vote has been found to be satisfactory.

[illegible]	Chairman: Homer R. Warner
[illegible]	[illegible]
5 -23- [illegible]	[illegible]

THE UNIVERSITY OF UTAH GRADUATE SCHOOL

FINAL READING APPROVAL

To the Graduate Council of The University of Utah:

I have read the thesis of James Carr Smith in its final form and have found that (1) its format, citations, and bibliographic style are consistent and acceptable; (2) its illustrative materials including figures, tables, and charts are in place; and (3) the final manuscript is satisfactory to the Supervisory Committee and is ready for submission to the Graduate School.

Date

Homer R. Warner
Member, Supervisory Committee

Approved for the Major Department

Homer R. Warner
Chairman/Dean

Approved for the Graduate Council

James L. Clayt n
Dean of The Graduate School

ABSTRACT

ILIAD is a microcomputer based program that has been developed to facilitate the construction and management of bibliographic knowledge bases. The contents of these knowledge bases are selected bibliographic citations to the literature. The knowledge base contents are organized by knowledge models that permit access to and the retrieval of relevant information. A knowledge structure referred to as a 'relation' is the fundamental unit of this knowledge model. It is shown that conceptual networks of relations may be built and utilized to perform high quality, efficient information queries of the knowledge base. Quality and efficiency are assessed by the measures of recall and precision, functions of the general relevance of information retrieved, and time.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to express my appreciation to several persons who have meaningfully contributed to the effort that is represented in this thesis. Firstly, to my advisor, Dr. Warner, who has willing shared his ideas, experiences and counsel, who conceived this project and helped realize its completion. Secondly to my committee advisors, Drs. Williams and Haug, for their time, interest, and input into the development of the ILIAD program and the achievement of its successful defense. Thirdly, To Dr. R.N. Hirst, who has constantly challenged me to think new thoughts and perform many faceted new tasks. I also gratefully acknowledge the contributions to my cause of many of the faculty of the Department of Medical Informatics. Their instruction, counseling, critiquing, and encouragement have greatly impacted the realization of this work and the continuing directions of my education. I also express my depth of gratitude to my parents who, as in all I do, have been my silent champions. Finally, to my sweet wife, for her love, patience, understanding and encouragement; tirelessly prodding me ever foward, words of acknowledgment are not enough. Thanks.

CHAPTER 1

INTRODUCTION

Amidst the current technological explosion and the overwhelming glut of medical literature in current publications, today's medical student and physician are faced with an enormous challenge of assimulating and organizing massive amounts of information.[1, 2] Responding to the ensuing need to develop effective strategies for dealing with this information overload, a concerted effort is being made to bring about computerized aids to information management.

The Need for Medical Information Management Skills

Reporting on the skills of medical information management, a subgroup of the AAMC Project Panel on the General Professional Education of Physicians and College Preparation for Medicine recommended that "medical facilities assess the extent to which their institutions offer opportunities for employing electronic data processing technology for fostering learning skills."[11] The need to provide opportunities for medical students to acquire knowledge about and skills in information management is increasingly recognized by the education community. [1,6,3,4,5,6,7,8] The AAMC panel further reported that it is skills in tracking down and assessing biomedical knowledge (most of which resides in the journals) that will more and more determine a clinicians effectiveness. These skills should be acquired during the preclinical training years and applied to clinical practice with refinement throughout one's medical career.

Clinicians face three challenges every day: (1) reaching the correct diagnosis, (2) selecting the management that does more good than harm, and (3) keeping up to date with advances in medicine.[9] In each of these clinical actions the necessity arises of using an approach to gathering and interpreting clinical evidence that will systematically facilitate the projection of diagnostic findings, prognoses, and therapeutic responses from groups of patients to individual patients in order to improve clinical performance. The continuing challenge of recognizing and responding to a need to change diagnostic and therapeutic maneuvers in order to remain consistent with valid new knowledge is central to the practice of medicine. One might conclude that all clinical actions and responsibilities reflect the clinician's priority of effectively managing and expanding their medical knowledge bases.[13]

The magnitude of the efforts required to meet these responsibilities may be illustrated by considering the magnitude of the rate of growth of the biomedical literature base. For example, to keep up with the 10 leading journals in internal medicine a clinician must read 200 articles and 70 editorials per month.[10] There are now over 20,000 different biomedical journals published (up from 14,000 10 years ago). To "read up" on viral hepatitis requires selection from among 16,000 citations published on this topic in English alone over the last 10 years.[12] The biomedical literature is expanding at a compound rate of 6% to 7% per year. Thus, it doubles every 10 to 15 years and increases 10-fold every 35 to 50 years. By contrast, the clinician's time available for the reading of the clinical literature is constantly being whittled away by other demands.

The Science of Medical Information Management

Applications of the science of medical informatics to information management revolve around the concepts of "data bases" and "knowledge bases." On-line bibliographic citation "data bases" contain indexed article citations to the published literature, whereas "knowledge bases" contain an analysis or synthesis of published information usually within a specific field. Knowledge bases may generally be thought of as "value added" databases. This is to say that in addition to the conventional organizational methods employed in data base management the contents of a knowledge base may also have been quality filtered or screened, more thoroughly indexed, and in many instances critically appraised by a medical expert or experts in order to ensure scientific validity . Bibliographic data bases are roughly analogous to printed indexes such as Index Medicus or to a library's card catalog. Knowledge bases are rough equivalents of textbooks or state-of-the-art reviews.[7]

Information Data Bases

A bibliographic data base contains references to the published literature and is most often used as a tool that guides one to a journal report. The effective utilization of these data bases requires that the user understand certain fundamental principles of automated information retrieval.

Searching bibliographic data bases requires the user to combine search terms in various ways using Boolean logic. Several search terms may be combined by using the Boolean set operators 'and', 'or', and 'not' so that sophisticated searches are devised. Analogous to printed indexes, bibliographic data bases rely on conventional human or machine-aided indexing with carefully controlled vocabularies (thesauri). Some data bases may also be searched by using words or parts of words that actually appear in the text. For instance, many automated data bases provide title, author, and source information while others also include a short abstract of the article or report as searchable fields. MEDLINE, for example, allows access through authors, index terms for a thesaurus, or words in the title or abstract.[11] A search can also be limited to certain publication years or languages. The searcher may additionally specify sex, age, or research with human or animal subjects. Complex on-line searches, of the MEDLINE type, are often performed by trained intermediaries.

The Institute for Scientific Information provides on-line access to its Science Citation Index and Social Science Citation Index through the Lockheed Information Systems' DIALOG system.[7] These two data bases, which have the same access and retrieval features as other conventional bibliographic data bases, can also be searched for authors and documents that cite or are cited by individual journal articles. This capability represents an important enhancement of subject-based retrieval, since it provides rich associative pathways among scientists and scientific research developments.

Although information data bases based on the bibliographic citation record are by far the most numerous, others are rapidly developing. Particularly useful are those that provide summaries of ongoing research projects and names and addresses of the principal investigators. For example, the NLM, in cooperation with the National Cancer Institute, maintains two such files related to cancer research: (1) CANCER-PROJ, containing approximately 16,000 summaries of ongoing cancer research projects in many countries and (2) CLINPROT, containing investigations of new anticancer agents and treatments.[7]

Knowledge Bases

The newest form of automated information retrieval systems is based not on data records per se, but on analyzed and organized "knowledge." Relevant information is selected, placed in a highly organized hierarchical arrangement to permit easy and effective retrieval, and encoded into a computer readable form. The contents of a knowledge base have therefore been analyzed or screened by the medical expert or experts who are directly responsible for the creation and up-dating of the "knowledge" content. Knowledge bases, by definition, are narrower in scope than data bases, dealing in general with only one area of medicine. This feature alone permits the knowledge base to lend itself to a much quicker mastery of any controlled key terminologies by a novice user.

Use of Automated Information Systems

Use of these on-line data and knowledge base systems is increasing and will undoubtedly continue as they become better known and are improved. However, the systems are not yet widely used. Even when used they are not often exploited to their fullest capacity. Part of the difficulty is that even though many of the systems do not require that the users be formally trained, most searches are performed by trained "search analysts." Such delegation of the search function is often inevitable given the complexity of and difference among existing services. The variability is a serious hindrance since it is unlikely that users will become familiar with all the files and systems they might need. Even trained searchers find it difficult to be fully conversant with several data bases and on-line systems.

Another major problem impeding the widespread use of automated information systems is the difficulty one encounters in selecting appropriate search terms while performing on-line searches. The searcher must think of all the possible ways to express a concept in anticipation of the words chosen by the author and indexer. An on-line dictionary file can be very useful in this respect, allowing synonyms for one concept to be pulled together. However, with no agreed upon standard for medical nomenclature, this process remains a major obstacle to the exploitation of the capabilities of computerized data bases.

MeSH is one of NLM's on-line, interactive dictionary files. It is an on-line thesaurus of medical terminology used in indexing journal citations for MEDLINE.[14] Thoroughly cross-referenced and coded to facilitate use of the MEDLINE data base, it is also, however, voluminous and unwieldy, impeding the novice user from benefiting from all its capabilities.

There are other barriers to widespread direct use of computerized information retrieval systems by physicians. Paradoxically, the superhuman speed and flexibility of computer searching often overwhelms the searcher with instant bibliographies of hundreds or sometimes thousands of citations, creating a formidable reading burden. Critical perusal may be made difficult by the lack of informative text and data in retrieved records. Moreover, it is likely that the search will not yield 100 % <u>recall</u> (the proportion of relevant documents retrieved from all the potentially relevant references) or provide 100 % <u>precision</u> (the proportion of relevant documents in the actually retrieved set of records). In general, precision and recall are inversely related and are rather elusive measures of operational retrieval systems because of the subjectivity of relevance judgments. Further frustration can result by not being able to locate articles listed in the printout.

Until computer aided retrieval systems surpass their conventional manual counterparts in overall ease, quality, and cost-effectiveness, many physicians will continue to rely on the more traditional means of gathering information.[7]

CHAPTER II

THE ILIAD PROGRAM

The structural component of knowledge and data base systems is the main focus of our research efforts. The structural component, as opposed to the content component, is that part of the system which is concerned with data organization and management. A data base that is effectively structured is easily accessed, both at the time of information storage and retrieval. In developing our data base management program, called ILIAD, the guiding paradigm has been to develop a program through which a medical expert could design and exercise direct control over this structural component. As will be shown this has been done through the use of knowledge modeling structures referred to as 'relations.' Relations are linked together by the program to form a 'conceptual network.'

The information management strategy employed within ILIAD is based on this concept of 'conceptual networks.' The conceptual network method is aimed at realizing of the following two objectives: (1) the minimization of the amount of time required of the expert in authoring the knowledge base; and (2) the optimization of the quality of the information retrieval process.

Definitions

This section presents the definitions of several concepts that are central to our conceptual network approach to information management. These concepts must be properly understood in order to benefit from the further discussion to be presented in this thesis.

'Knowledge representation' refers to the format in which the content component of the knowledge base is stored within the computer. The knowledge bases managed by the ILIAD program use the form of bibliographic citations to journal articles as a format of knowledge representation. These citations represent journal articles that have been selected by the expert according to their relevance , currency, and validity. By design the expert assumes direct control over this component of the knowledge base, ensuring the selection of a relevant, high quality subset of the literature.

'Keyword' refers to any word or term that is used in indexing citations. Keywords may actually be a combination of two or more words that together express a concept (e.g., Artificial Intelligence). 'Concepts' are therefore represented by the combination of one or more keywords. A 'relation' is a label that is given to single hierarchical unit or node in the conceptual network. A relation consists of a stem concept that branches to one or more keywords. These branches define features or properties of the stem concept. The 'conceptual network of relations' is constructed automatically by the ILAID program by linking relations together. These links are produced by identifying nodes in the network that have the same keyword as a stem or branch.

'User' refers to those persons who use the ILIAD program to gain access to an already existing knowledge base. 'Expert' refers to those medical experts responsible for the design of the knowledge structures or models and the selection of the knowledge base content.

Program Operation

The main functions of our data base management program are information storage and retrieval. The major features of the ILIAD program associated with these functions will be presented via the following scenario of program operation.

At run time the ILIAD program presents the user with a menu of options that direct program operation (see Main_Menu, Appendix A). A "menu" is a list of program options displayed on the computer screen. Each of these options is identified by a unique alpha-numeric character. The user may then select the desired option when responding to the "CHOOSE OPTION" prompt by selecting the desired unique identifier from the menu list. This usually entails a single keystroke followed by a carriage return.

At the time of information storage, an "A" is selected to choose the "A)DD CITATION" option. The operator is then prompted, through a line-oriented interface, to enter standard bibliographic information for citation identification (see Table 1). The length of each field, in terms of the number of characters permitted by the program is displayed graphically on the computer screen at all times during data entry. The operator is thereby constantly reminded of the length of the text string that will be accepted by the program.

The program requires that journal names or other citation sources be entered in abbreviated form to ensure a unique representation of every source within the limits of 20 alpha-numeric characters. If the operator attempts to enter a journal abbreviation that has not been defined previously an alert is given of a possible data enty or typographical error. A list of acceptable journal abbreviations is then displayed in menu driven format so that the operator may select the appropriate abbreviation. A new journal name and abbreviation may be added to the data base in the case that it has not previously been referenced. The Table of Abbreviations, as it is called, is maintained in alphabetical order so that it may be perused by an operator who is unfamiliar with a particular journal's abbreviation. This option is invoked by a single keystroke as are all other program options.

The program retains in memory the bibliographic information entered during the citation addition process. In instances where several articles are entered consecutively from the same journal issue or volume only that data which are unique to the citation information of a particular reference needs be entered manually. The remaining data are entered automatically by the program. After the data entry portion of citation addition is completed the citation is assigned a unique accession number. These numbers are assigned to each citation, beginning with 1, in a sequential order.

After bibliographic data entry is completed the citation is then displayed in standard format on the computer screen. Before these data are actually stored to disk the operator is instructed to proof read the data and edit the citation if spelling or typographical or other errors are detected.

The indexing process, assigning keywords to a citation for its later retrieval, then begins. The date, journal abbreviation, and authors' names are first entered automatically by the program as keywords. The article's title is passed through a filter to remove stop-words, i.e., prepositions,

Table 1

Fields and their Descriptions for the Article Citation Record

FIELD NAME	DESCRIPTION
TITLE	[Non-"stop-words" are Searchable]
AUTHORS	[UP TO 8 AUTHORS]
JOURNAL	[Using the correct abbreviation]
DATE	[or Publisher for Book Citations]
VOLUME	[or Place of Publication for Books]
ISSUE	
PAGES	
ABSTRACT	[Optional Narrative Field, Not Searchable]
SEE ALSO:	[Cross-referencing field]
MODIFIERS	[Defined By the Expert]
KEYWORDS	[Up to 20 keywords per citation]

adjectives, etc., from consideration as keywords and the remaining title words are entered as keywords. The expert author may have defined a Topical Table of Contents or a Table of Modifiers previous to this time of citation addition. These knowledge structures would then be invoked automatically, based on keywords from the article's title, to facilitate the citation indexing process. Keywords from the title that match to any heading terms in either of these tables would cause appropriate subheadings to be displayed. The operator selects from these branches any and all terms that correctly describe the article in more specific detail than the broader heading term. The program will continue to automatically prompt the operator to be more and more specific until the desired level of detailed description has been specified.

These knowledge structures may likewise be invoked manually by the operator. The operator may also enter additional keywords that do not appear in the citation information or in the tables. These are keywords that further characterize the article's content. This entire citation addition process requires an average of 3 to 5 minutes.

To facilitate the information retrieval function the ILIAD program allows the expert to define a rich conceptual network wherein concepts are defined and linked to associated concepts. This network is constructed of nodes referred to as 'relations.' The relation knowledge structure consists of a stem keyword and keywords that are labeled branches to the stem. The stem keyword may represent a concept or a category of information. Its branches then represent features or properties of the stem. For example, if the category represented by the keyword 'animal' were the stem of a relation, its branches might include 'dog', 'cat', 'horse', 'cow', etc. If the set of all automobiles represented by the keyword 'cars' was the stem of a relation, its branch items, forming a subset of the set of all cars, might include 'Ford', 'Chevrolet', 'Oldsmobile',etc. Furthermore, the keyword 'Ford' may in turn be the stem of a relation. In which case the branches might include types of Ford cars, i.e., 'Pinto', 'Mustange', 'Thunderbird', etc.

The use of the conceptual network in searching may broaden the scope of a search. By using the conceptual network to automatically pull together lists of related keywords and concepts the searcher is not constrained by the requirement of specifying the exact keywords that were used by the person responsible for storing the citation record. By entering more than one keyword or by combining the stems of more than one relation the scope of the search may again be narrowed. This is done, however, without as high a risk of missing the desired item since each keyword has been expanded to include 'related' terms through the conceptual network.

Returning to our scenario of program operation, entering an "S" to select the "SEARCHER" option from the Main_menu invokes the program's searching algorithms. These algorithms will automatically use the conceptual network as previously defined by the expert to prompt the user in the selection of appropriate search terms.

The user may also interact with the tables used in indexing the citations in order to formulate the search criteria. These tables may be made interactive so as to guide the user with screens of prompts in order to effectively utilize the structure of the knowledge base. If, however, a user is familiar with the underlying structural framework of the data base to the extent that this prompting mechanism would be tedious to follow the Topical and Modifier Tables of Content prompts may be bypassed.

In either case the method is followed of systematically, and automatically moving from general to more specific, detailed descriptions of search parameters utilizing the knowledge created by the expert. This process relieves the user of the need to be concerned with Boolean logic in the search strategy formulation. One or more general concepts may in this way be defined in sufficient detail and then combined in order to retrieve a specific and highly relevant set of citations. If a given search strategy proves to be too narrow or too broad the process may then be repeated.

A typical scenario would find the student beginning an information query by entering keywords that express concepts to investigate. Associated keywords are then identified and presented for the student's consideration in refining a search strategy through the conceptual network built by the expert. By refining the search strategy it is meant that the student may broaden the scope of the search by including related concepts or narrow the scope by selecting more specific branches to more strictly define a concept. Additionally, synonyms of keywords entered, as also defined by the expert author of the knowledge base, are automatically included in the search criteria for retrieving all related citations. The student would then be presented with a list of citations retrieved as matching the search criteria. At this time these citations may be reviewed one at a time. If the user desires then to retrieve the reprint for further study a hardcopy of the bibliographic data may be generated along with its accession number for locating the document in a reprint file. In addition, based on a review of specific keywords used to index retrieved records, another search strategy may be formulated in order to retrieve other related citations.

To illustrate in a simplistic way the use of this type of conceptual network in logically organizing and retrieving knowledge consider the following example. Suppose that an article in the National Geographic entitled "The Tuna Industry off the Coast of Nova Scotia" is only vaguely recalled. The searcher is not able to remember the exact geographical location in Canada nor the exact type of fishing reported. A search could be performed using both of the keywords 'Fish' 'and' 'Canada' that would find and retrieve the citation given the relations represented schematically in Figure 1. After the user is satisfied of having retrieved an acceptable set of relevant references the program may be terminated by simply selecting the "X," for "EXIT," option from the Main_Menu.

Additional Features of the ILIAD Program

An additional aspect of the concept definition capability provided by ILIAD is a keyword dictionary which allows an expert to define words that are considered synonomous of a given keyword or phrase. For example 'ECG' may be assigned the synonyms of 'EKG' and 'electrocardiogram'.

Another type of relationship that the expert may define between concepts permits the cross-referencing of related topics to each other. These may be cross-references to keywords within the same knowledge base or cross-references between separate but related bases. The computer program manages these predefined cross-references or "see_also" associations by

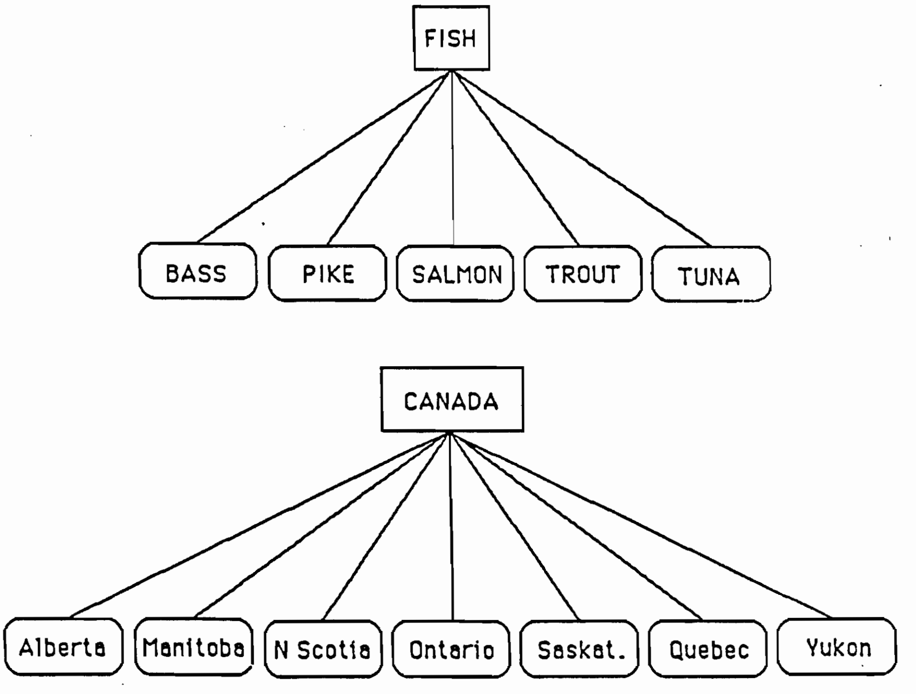

Figure 1: Examples of Simple Relations used as part of a Search Strategy

automatically identifying and displaying topical cross-references for the user's consideration in further refining search criteria.

An important aspect of ILIAD's knowledge structures is a potential for them to be changed dynamically with time. To be dynamic means that the model needs to be able to evolve with time and adapt to new ways of viewing medical topics in light of advances in medicine or the continuing learning processes of the expert. Knowledge base management tools are thereby provided within ILIAD to accommodate this evolutionary behavior.

ILIAD features an option refered to as the "Analyzer" which permits a user to automatically analyze the knowledge base in order to identify the context within which any given keyword is used. Identifying context dependencies provides information for the searcher's consideration in order to correctly formulate search criteria. The ultimate objective is to help searchers deal with the inherent variability and ambiguity of natural language.

ILIAD provides a facility refered to as the "Completer." The Completer permits the clever user to explore the natural usage of keywords in article titles. By entering a truncated portion of a keyword in question the Completer will automatically scan the knowledge base and identify and display all possible "completions," words with the same root portion but different suffixes (e.g., ing, ed, s, es, etc.). The user will then be prompted to consider all possible variations of the root term, including hyphenated terms, and may select from them any and all terms with the same apparent meaning (see Figure 2).

Our development efforts have focused on (1) terminology selection to complement natural language usage through the use of synonyms defined in a data dictionary, (2) retrieval protocols requiring no knowledge of boolean logic through the use of the TTOC and relations, and (3) friendly user interface protocols using menu driven options for program operation.

Program Implementation

The program features described above have been implemented using the programming data structures and software tools outlined below. File designs and dimensions pertaining to the current implementation are also outlined below. Our most current ILIAD implementation is on a

Entering the term "CARDI" returns:

CARDIAC
CARDIOLOGY
CARDIOPULMONARY
CARDIOVASCULAR
etc.

Entering the term "SIMULAT" returns:

SIMULATE
SIMULATED
SIMULATING
SIMULATION
SIMULATOR

Entering the term "ACID" returns:

ACID
ACID-BASE
ACIDS

Figure 2: Examples of the Completer. Any or all of these terms may then be selected at the time of search strategy formation, citation indexing, and building Relations and the TTOC.

LEADING EDGE PC, a semiclone of the IBM XT. The system is configured with a 20 megabyte hard disk system. This machine was targeted as the development type because of its computing power, secondary storage capacity, and wide spread availability throughout the health sciences community. The software for the system has been developed using the C language because of its relative easy portability to other machines.

Data Structures

The major functions of the ILIAD program revolve around the storage and retrieval of information. The following data structures are used to build the knowledge models that facilitate these functions.

Topical Table of Contents. (TTOC) Much like the table of contents of any textbook, a TTOC is developed by the expert as a superstructure or framework within which the knowledge base is organized. The main headings of the TTOC structure may be classified according to topics selected by the expert, reflecting personal preferences, e.g., organ or body systems, symptoms or signs, clinical problems, specific target disorders, anatomic sites, psychosocial concepts, physiological concepts, various technologies, etc. The process of developing the TTOC is then one of continually defining sublevels of headings under each mainheading in the table so as to be as specific or precise in breaking down a given topic as need be. Up to five levels of detail may be developed within ILIAD's current dimensions. As one moves down the levels of the TTOC one moves from general to specific classification information (see Appendix B).

Index. Every keyword entered into the data base for the indexing of a reference is automatically inserted into an alphabetical index of keywords. Keywords are entered at the time of bibliographic citation data entry. This index also contains information regarding the numbers of citations indexed by each keyword and the usage of the keyword as the stem of a relation.

Relation. Relations are defined by the expert to link related keywords and concepts together. A keyword is first identified as the relation's "stem." The relation may then be used to define three types of associations. (1) Synonyms of the stem may be defined that form logical and physical linkages between keywords for retrieval purposes. Performing a search that uses the stem as part of the criteria for retrieval will automatically include keywords that are identified as its synonyms with a Boolean 'or' operator. (2) Another set of related keywords may be defined as a "subset of" or "type of" the relation's trunk. These keywords identify features or properties of the stem and are labeled "branches of the stem." (3) Linkages between a stem keyword and its branches may also reflect an association between concepts that reflects the expert's understanding of how concepts are related to each other (e.g., certain diseases may have related complications, or similarly may cause other related diseases).

The relationship between a stem and its branches is used to facilitate the formulation of search criteria much like the intended use of synonyms. When a relation's stem is used as a part of the search criteria its branches are automatically included. Each of the keywords that are defined in a relation as branches is therefore also associated together for searching purposes by the logical Boolean 'or' set operator. A relation's stem is used as a search criterion when the scope of a search is to be broadened or generalized. Branches may be selectively chosen as more precise means of expressing the concepts of interest in order to narrow the scope of a search.

Citation. The Citation structure is used for storing in standard form a bibliographic citation to the literature, e.g., a journal article, a medical textbook, etc. The expert's clerical staff is prompted to enter all required citation information interactively by the program (see Table 1 and Figure 3).

Modifiers. A Modifier is used to further characterize or classify a cited reference. Its usage is much the same as the TTOC for indexing purposes. The expert may create a table of modifiers that are general descriptors of journal articles. By definition, Modifiers traverse the topical classification scheme outlined by the TTOC and contain descriptors of articles that apply to any topic (e.g., reprint location, article type, study type, etc. See Figure 4).

Abbreviations Table. A table of journal abbreviations is created by the expert to ensure consistancy over time with the manner in which journal names are abbreviated. During citation entry only abbreviations found in the Table of Abbreviations will be accepted as valid. The

#253

COMPUTER ANALYSIS OF SERIAL ELECTROCARDIOGRAMS

PRYOR TA, LINDSAY AE, ENGLAND RW

CBR, 1972 DEC;5(6):709-714

<CBR >	<1972 DEC >	<ANALYSIS >
<SERIAL >	<ELECTROCARDIOGRAMS>	<PRYOR TA >
<LINDSAY AE >	<ENGLAND RW >	

---FORMAT EXPLANATIONS--

Accession Number

Title

Authors

Journal Abbreviation, Date;Volume(Issue):Pages

<Keywords...>

Figure 3: Example of a Citation in Standard Bibliographic Notation with Format Explanations

1. LOCATION
    - REPRINT FILE
    - STACKS
    - TEXTBOOK
    - ECCLES LIBRARY
2. MeSH SUBHEADINGS
    - ANATOMY & PATHOLOGY
    - DIAGNOSIS
    - EDUCATION
    - ETIOLOGY
    - METHODS
    - THERAPEUTIC USE
    - etc.
3. ARTICLE TYPE
    - ORIGINAL ARTICLE
    - REVIEW ARTICLE
    - EDITORIAL
    - LETTER TO EDITOR
    - ANIMAL STUDY
    - etc.
4. TYPE OF STUDY
    - CASE CONTROL STUDY
    - COHORT STUDY
        - RETROSPECTIVE
        - PROSPECTIVE
    - CROSS SECTIONAL (PREVELANCE
    - CLINICAL TRIAL
5. REASON FOR READING
    - DIAGNOSTIC TEST
    - CLINICAL COURSE
    - ETIOLOGY/CAUSATION
    - THERAPY APPRAISAL
    - COST EFFECTIVENESS
    - PROFESSIONAL NEWS
    - CRITICAL THINKING
    - etc.
6. CRITICAL APPRAISAL
    - CLINICAL(THERAPEUTIC) UTILITY
    - EPIDEMIOLOGICAL VALIDITY
    - BIOSTATISTICAL VALIDITY
    - etc.

Figure 4: Examples of a Modifier Table for Article Classification

program therefore forces the usage of abbreviations predefined in the Table of Abbreviations . When a citation's source is entered into the knowledge base as part of the citation data that does not match to any previously defined abbreviation, the user is alerted of a possible error. Incorrect abbreviations or spellings may then be identified and corrected before the citation is stored to disk. The user also has the option of creating a new abbreviation in the Table of Abbreviations if the source has not yet been cited. If the document's source is not a journal the abbreviation control algorithm may be overridden.

Software Tools

Sevaral general software tools or algorithms have also been developed to facilitate the citation storage and retrieval functions.

Indexer. The Indexer is invoked when a new citation is added to the data base for the purpose of indexing the citation using keywords. Certain keywords are entered automatically as indexes to the reference. They include the journal abbreviation, date of publication, modifiers, and TTOC headings selected by the expert. The program interacts with the operator by using the TTOC as a prompting mechanism for indexing the reference within the appropriate knowledge structure and classification schemes. Author's names and title words, that are not filtered out by a subroutine which eliminates stop-words, may be entered directly at the discretion of the operator. In other words if a title keyword does not convey meaningful information in regards to the content of the article, it may be deleted from the list of index keywords. Every keyword used for indexing is mapped to a unique numeric code by a routine that controls data definition. The coding mechanism is used to improve the response time of the system during citation retrieval and to improve the accuracy with which the searching procedure is performed.

Searcher. This tool is used for the definition of search criteria. The user selects from an option's menu the choice of searching via the TTOC or via the Index. If the "TTOC option" is selected the TTOC prompts are presented to the user. These same prompts were supplied to the expert or clerical staff operator at the time of citation entry and indexing. As headings are chosen from the mainheadings list lower levels of subheadings are automatically identified and presented to the user for consideration in refining the search criteria. With each subheading chosen lower levels of subheadings are presented in a like manner until the retrieval criteria are fully defined to the user's satisfaction or until all options in refining the criteria have been exhausted. This type of searching is analogous to looking up keyterms in a medical textbook via the table of contents. Each keyword selected is limited by the context of its usage within the topical structure of the TTOC, e.g., "treatment" may be limited to "treatment of Rheumatoid Arthritis. " The user may further narrow the scope of the search by entering additional keywords not found in the TTOC, e.g., date of publication, author, journal name, Modifiers, etc. Once the criteria have been defined, the search is performed. The number of citations retrieved is then displayed to the user. The choice may then be made to review each of the citations retrieved with an option to generate a printed hardcopy in standard bibliographic notation, or to repeat the searching process after making modifications to the retrieval criteria.

As opposed to the topic oriented search performed by the "TTOC search" option, the "Index search" option may be used to search across medical topics. The search criteria are defined by entering keywords in freetext format. These keywords are then mapped to terms in the index through the data definition routine. An Index search is analogous to looking up the location of keyword concepts in a medical textbook via the index. It may hence be completely independent of any topical context. Searching with the index option causes relations from the conceptual network to be invoked as keywords match to the stem or branch words of the relation. The user is then automatically prompted to consider related keywords in expanding or refining the search criteria.

Completer. The Completer takes an alpha-numeric string entered by the user, a "root" term, and then scans the alphabetic index for strings that could be considered "completions" of the root term entered. The rule for selecting completion words states that possible matches are identified by selecting every string that might be truncated to match the root. Case is not considered by the Completer. A list of potential matches is then presented to the user for selection of the string or strings to be used by the Matcher or Indexer. The Completer may be used with ingenuity to define optimal search criteria when uncertainty exists as to the precise usage of terminology by the knowledge base's author, the expert, or the article's author. Ingenuity is rewarded when selecting search terms interactively with the Completer in the Index search mode. For example, the root term "cardi" entered in freetext format may be used to retrieve citations indexed with both the keywordss "cardiac" and "cardiovascular" (see Figure 2). The Completer may be used when the spelling of the author's name is not known with certainty or when middle initials are not used consistently. Entering only a portion of the last name would cause the completer to identify all possible matches in the index and display them to the user, permitting the selection of one or more variations of the same name.

Matcher. The Matcher is the algorithm that executes a search request. It first checks to see if keywords entered have any predefined associations or linkages. In general, all keywords entered are logically related by the Boolean 'and' set operator. Two exceptions to this rule are synonyms and branches of a relation. These terms are considered logical 'or' extensions of a search criterion. Regardless of the particular Boolean relationship, all Boolean operations are managed by the Matcher in a manner that is transparent to the user. The Matcher uses special purpose data base management tools to perform the search and returns pointers to each citation found as satisfying the search criteria.

Analyzer. The Analyzer is a tool designed for aiding in the definition of search criteria. When a keyword is "analyzed" all instances of its usage in the TTOC, Modifier Table, and conceptual network are located and displayed to the user. The recall and precision of a search strategy may be increased if the user takes time to analyze the knowledge base before the search is performed and identify the context within which keywords have been defined and used in the knowledge structures, i.e., relevant citations will not necessarily be retrieved if keywords are used out of their proper context in the TTOC. The analyzer will also show any associations or linkages to the keyword in question. Synonyms, branches and cross-references are therefore displayed for the user's consideration in formulating the search strategy.

Browser. The Browser is also a tool used for aiding in the development of the search strategy. It may be used to browse through the TTOC or through the alphabetical INDEX of keywords or through the Modifier table. Part of its function is to display a count of the number of references indexed by each keyword "browsed." This allows the user to approximate the scope of a search criterion prior to its actual use. It also provides a useful tool for the expert's use in the management of the knowledge base.

Knowledge Base Management Tools

Editor. This is a group of editing tools to be used in editing any and all instances of the knowledge base data structures.

Displayer. The Displayer allows any and all instances of a data structure to be sent to a printer for the generation of a hard copy of the structure. It may also be used to display the data structures to the computer terminal for review.

Up Dater. This tool is used for the managment of the keywords assigned for the indexing of references and allows for the dynamic management of the knowledge base and the knowledge models as they evolve with time.

File Design

ILIAD's customized database management system, DBMS, manages 34 separate data base files stored on a hard disk system. A set of Indexed Sequential Files is used so that access may be either sequential or direct into any portion of the data base. At the time of program start-up files containing the data structures described above are read sequentially and pointer or index files are built to provide direct, random access to any given instance of a data structure stored on disk. Disk I/O is managed by two subroutines that control all file access when both reading from and writing to the data base files. This design was considered because of a desire to maintain a high level of software portability. Machine dependencies that govern I/O need only be dealt with within these two subroutines.

Thirty-four separate files are used in order to minimize both wasted space on the storage medium and seek time, the time required to locate a specific data record in the data base. The file structures themselves are based on the data record structures outlined in Tables 2 - 7. There are separate files for each of the main data record types. These include: (1) a citation record file, (2) a TTOC and relation record file, (3) an abbreviation record file, (4) a modifier record file, and (5) a definition record file. The file records have been designed so that space would be allocated only for data that are actually defined. If, for example, a term is not assigned a definition or a See_Also field then no disk space is allocated for these fields. To accomplish this design criterion a system of pointers forming linked lists of records is utilized. A pointer that is set equal to zero indicates that the corresponding field has not been defined and hence no storage space allocated. Twenty-seven of the 34 data files are used by the program's alphabetical index of keywords, i.e., one file for each letter of the alphabet and one file for miscellaneous numeric keywords. This arrangement provides for an index that is large enough to be functional given the limitations of internal memory accessible by the ILIAD program. These index files are read into memory only as they are needed.

A header file contains information necessary for the management of and access to the entire data base file system. This file is the first file that is read when the program is initiated in order to provide access information for the DBMS in managing the file system. The header file also contains information regarding passwords so that measures may be taken if desired in order to provide a degree of security for the data base. This password security system may be used to protect to a certain measure the integrity of the knowledge base.

The Medical Informatics' prototype knowledge base requires approximately 900 kilobytes of data storage. The dictionary file itself requires 620 kilobytes and the citation data file requires 210 kilobytes. The knowledge base contains 824 citations and the knowledge structures, tables and relations as listed in Appendix B. The 27 combined index files require a total of approximately 70 kilobytes of data space.

TABLE 2

Citation Record Structure

RECORD NAME	FIELD NAME	DIMENSIONS
CITATION		
	TITLE	160 characters
	AUTHORS	80 characters
	JOURNAL	20 characters
	DATE	12 characters
	VOLUME	5 characters
	ISSUE	5 characters
	PAGES	10 characters
	SEE_ALSO	40 characters
	KEYWORDS	20 characters * 20 keyterms/citation

LIMITS: No actual limit.

TABLE 3

TTOC Record Structure

RECORD NAME	FIELD NAME	DIMENSIONS	
TTOC			
	MAINHEADINGS	40	Headings
	SUBHEADING	4	levels * 20 Headings

LIMITS: 400 unique keywords as headings and subheadings in the TTOC

TABLE 4

Relation Record Structure

RECORD NAME	FIELD NAME	DIMENSIONS
RELATION		
	TRUNK	20 characters
	SYNONYMNS	4 synonyms * 20 characters
	BRANCHES	20 branches * 20 characters

LIMITS: 100 relations per Knowledge Base

TABLE 5

Index Record Structure

RECORD NAME	FIELD NAME	DIMENSIONS
INDEX		
	KEYWORD	20 characters
	DEFINITION	50 characters
	SEE_ALSO	40 characters

LIMITS: 300 keywords per letter of the alphabet = 7800 unique keywords
LIMITS: 2000 instances of each keyword for indexing purposes = 15,600,000 keywords possible for indexing citations

TABLE 6

Abbreviation Record Structure

RECORD NAME	FIELD NAME	DIMENSIONS
ABBREVIATION		
	ABBREVIATION	20 Characters
	UNABBREVIATED TEXT	30 Characters

LIMITS: 100 Abbreviations in the Table of Abbreviations

TABLE 7

Modifier Record Structure

RECORD NAME	FIELD NAME	DIMENSIONS	
MODIFIER			
	MAIN MOD.	20	modifiers * 20 characters
	SUB MOD.	20	submodifiers * 20 characters

LIMITS: No limitation on the number of submodifier levels as with TTOC

Program and File Dimensions

Current versions of ILIAD are capable of defining 40 different mainheadings within the TTOC. Twenty-seven separate data bases corresponding to different medical disciplines may be defined by the expert each having its own, separate TTOC. Approximately 8000 unique keywords and abbreviations may be defined in ILIAD's index . This is the most limiting dimension of the data base program. Each of these keywords may be referenced up to 2000 times. Each of the 27 separate data bases has its own index with these dimensions.

The Topical Table of Contents in each instance can have up to as many as four levels of depth defined below each of its 40 main headings. Each of these four sublevels in the table may in turn contain a list of up to 20 headings (see Appendix B) . These headings as with all other keywords used by the program are each mapped to the data dictionary by the data definition language where they are assigned a unique numeric code for manipulation by the program's algorithms.

The program allows 20 alpha-numeric characters for the expression of each of the keywords in the TTOC and the INDEX. These 20 characters must form a unique representation of the term or concept. If a 20 character representation is too limiting for adequate expression of a medical concept the data dictionary permits further definition and explanation capabilities. A field of an additional 50 characters is available for this more elaborate definition. Definitions are displayed to the user whenever the keyword is used in one of the program's data structures or algorithms. The data dictionary further permits the expert to define a See_Also field of 30 characters as a means of prompting the user to mentally associate cross-reference links between keywords or concepts. These See_also, cross-reference, associations implicitly link concepts together according to the expert's understanding of the contents of the knowledge base.

The Relation structure has dimensions similar to those of the TTOC. A stem may have up to four synonyms defined and a list of up to 20 branches. Branches of one relation may then be stems of other relations, to any depth of nesting (see Appendix C).

Most of the limitations imposed on the data base size are artificial and totally arbitrary (see Tables 2 - 7). These were chosen as dimensions to satisfy the criterion that they be sufficiently large so as to not restrict the development of the concept of a personal knowledge base of bibliographic data. The only criterion for their selection was to insure the program's functionality. They may therefore be enlarged as customized applications may demand.

The Matcher algorithm will manage up to 10 keywords combined using the Boolean 'and' set operator. In addition each of these 10 terms may be combined with up to 120 keywords using the Boolean 'or' set operator. This is designed to accommodate the usage of synonyms, branches of relations and other predefined or temporary linkages between related keywords.

Finally, the number of citations that it is possible to manage within one knowledge base system has no maximum . A practical number will be determined by experience, with the limiting factor being that of the effect of the size of the files on the speed of the program's execution. The number of unique keywords possible may also prove to be a limiting influence on the number of citations possible.

Building Knowledge Bases

Several prototype knowledge bases have to this date been developed using ILIAD, the most extensive of which is the MEDICAL INFORMATICS knowledge base, developed under the direction of Dr. Homer R. Warner. This is the knowledge base that has been used for the experimental purposes of the specific study to be reported in this thesis. From our experience with prototype knowledge base development projects we have identified several general principles that may be used to guide the development process of subsequent knowledge bases.

The building of a knowledge base consists of several tasks that may generally be separated into two main catagories. These are (1) Clerical tasks and (2) Knowledge Engineering tasks (see Table 8). While these tasks are closely related in function the temporal separation of their execution could expedite the knowledge base development phase. While it is felt that the expert should generally direct the knowledge engineering tasks, the time consuming clerical tasks may be performed by clerical staff independent of the expert's intervention.

Table 8

Knowledge Engineering and Clerical Tasks

Involved in Building Knowledge Bases

Clerical Tasks	Knowledge Engineering Tasks
1. Citation Data Entry	1. Defining Cross-references
2. Citation Indexing	2. Defining Keywords
3. Data Base Management	3. Defining Linkages (Synonyms)
4. Reprint File Management	4. Building TTOC
5. Library Management	5. Building Relations
	(Used to construct the conceptual network)

The processes requiring the expert's involvement in the authoring of a knowledge base containing bibliographic information may be described as involving the following knowledge engineering tasks. (1) The expert first selects an information source, e.g., a journal article, that is judged to contain relevant information, and has a staff member enter its bibliographic data into the system. (2) The expert may then identify keywords, descriptive of the articles content, for keyword indexing of the reference. These keywords are entered by clerical staff members as part of the citation information record maintained within the data base. This step is facilitated by the utilization of portions of the Tables of Contents that may already be developed. (3) The expert may subsequently generate any appropriate new "relations," "linkages," and "associations" between keywords and concepts to be included as extensions of the knowledge bases' conceptual network. In other words, as the relationship of an article being entered to other articles already in the system is considered, a 'relation' may be thought of that might facilitate the later retrieval of this and other related citations. For example, if a keyword for an article being entered was 'joint', the association between of 'joint' and 'arthritis' might be created so that any time thereafter that the term 'arthritis' was entered within a search statement this article's citation along with any others whose keywords included 'arthritis' 'or' 'joint' would be treated as 'hits' and retrieved as having satisfied the search criteria .

In certain circumstances it may be useful to enter articles that have not yet been carefully appraised and subjected to the scrutiny of careful reading. Such may be the case if a physician does not now have time to critically review an article but also does not wish to accumulate a "gonna pile" (an accumulation of journal articles you feel you "have to read") . In such instances the citation could be flagged as "unread" when entered into the data base. A user would thereby be reminded that the reference has not yet been critically appraised and that such appraisal is necessary before the validity, reliability, credibility, utility, or applicability of the findings reported in the article may be assumed.

CHAPTER III

METHODS

Developing methods for structuring knowledge so that it is highly organized within a set framework in the data base has been the primary activity of this research project. These methods require structures that facilitate the storing of information into the data base and the subsequent retrieval of information from the data base. The knowledge structuring methodology shall first be discussed. A brief description of the searching methodology will then be presented. This will be followed by a presentation of a specific study we have performed in order to validate the effectiveness of these knowledge structuring tools as a means of managing a knowledge base.

Knowledge Structuring Mechanisms

There are two structures that shall be further discussed in this section; (1) the conceptual network, used to facilitate the retrieval of information from the knowledge base, and (2) the Table of Contents (TOC), used to accommodate the storage of new information into the knowledge base for later retrieval.

The conceptual network is conceptualized as a hierarchical tree of keywords. As one traverses the tree from top to bottom the concepts expressed by the keywords go from general to specific. In addition to this hierarchical arrangement of keywords, additional links between keywords within nodes at different levels in the tree are established, thus creating a true hierarchical network. These network links are made be entering identical keywords as either stems or branches of a node in the tree. The conceptual network is managed so that access to any level may be gained by entering an appropriate keyword, stem or branch. The possibilities are then presented of moving down the network to more specific concepts, moving up the network to more general concepts, or traversing the network to a related but different concept. This is an interactive process that is presented to a user when querying the system to retrieve information. Each keyword in the network may additionally be linked to one or more synonyms. This provides multiple keyword points of access into the network. The knowledge base maintains two sets of inverted files for management of keywords. One file maintains pointers to citation records that are indexed by the keyword. The other file maintains pointers to nodes in the conceptual network and is used to manage the applications which use the tree for information organization and retrieval.

As stated the basic unit for the construction of the conceptual network is a knowledge structure referred to as a 'relation', a 'node' in the network. A relation consists of a central concept, 'stem', that branches to related keywords that are labeled 'branches.' As an example, the stem 'rhythm' may be linked to types of rhythms such as 'fibrillation', 'tachycardia', 'flutter', etc. Relations are hierarchically networked together in order to fully express the knowledge organizational schemes. In addition, nodes in the network may be linked to concepts in the Table of Contents to integrate both of these knowledge management structures. Relations are presented to the computer as separate entities which the program then automatically links into a conceptual network by mapping identical keywords to each other. The addition of relations to the network is, therefore, a relatively simple task and may be performed at any time after the data base is built and indexed.

The time required for the expert to develop this knowledge model is independent of the size of the data base. Furthermore, the model may be built after the clerical tasks are completed. Relations between keywords and concepts need not be anticipated by the person storing the data. The relation knowledge structure may also be used to classify the same citation record according to different classification schemes, e.g., functional versus anatomical classification of organ systems. The knowledge structure may be easily modified, with the program managing the actual construction and modification of the conceptual network. The salient characteristics of the conceptual network of relations as a tool for retrieving information are summarized in Table 9.

Table 9

Features of the Conceptual Network of Relations used in Information Retrieval

1. The time required of the expert to develop this knowledge model is independent of the size of the data base.

2. The model may be built after the clerical tasks are completed. Relations between keywords and concepts need not be anticipated by the person storing the data

3. Multiple relations may be built to retrieve the same information based on different classification schemes. e.g., functional versus anatomical classification of organ systems. Likewise, a term may be part of more than one relation, thus allowing the data base to express a variety of heirarchial structures.

4. The knowledge model may be easily modified.

5. Relations permit the use of Boolean OR logic only.

6. The conceptual network of relations is interactive at retrieval time to prompt the user in considering related concepts as extensions of the search criteria as well as considering branches that are a subset of a given concept.

7. An expert in a given field can provide the linkages between and organization of concepts within the knowledge base to facilitate the successful querying of the system by a nonexpert.

To facilitate the organizational tasks of managing the content component of the knowledge base, especially when storing new data to the data base, we have developed a knowledge structure referred to as the Table of Contents (TOC). This table is designed and built by the expert as a means of defining a general framework for the structuring of the data base content. Within this table, analogous to the table of contents of any textbook, concepts are implicitly defined as represented by keywords. By interpreting keywords in view of their position in the table, or context, one may reliably ascertain their conceptual meaning. In this way the keyword 'language' may be differentiated from the context of 'computer languages to be executed' and 'natural languages to be processed.'

When a new citation is entered into the data base the expert indexes the article according to one or more schemes that correctly classify its content. This is done by correctly "pigeonholing" it in one or more locations within the TOC framework. The mainheadings in the Table of Contents are analogous to the chapters of a textbook. Each of these chapters may in turn have sections defined within it ('subheadings' in the TOC). Each of these sections may be devided into subsections and so on. The 'chapters' in the TOC are not hierarchically organized, so that new areas of interest may easily be added to the table. There may be chapters on as diverse and broad a subject matter as needed in order to comprehensively classify the content, e.g., body systems, diseases, physiological concepts, signs or symptoms, etc. Related concepts within each of these chapters may then be linked together with a relation.

The Table of Contents permits the usage of strings of words for the expression of concepts (e.g., Information Systems). In this way the expert may create a quasi-controlled vocabulary environment for the indexing of citations. The TOC also permits the usage of both Boolean 'and' and 'or' logic when combining keywords to express a topic.

In order to fully benefit from the capabilities of the TOC the expert must review each record one at a time and manually assign them to positions within the TOC. The data base is then organized within this framework during data entry. Any time this knowledge structure is updated the contents of the data base must be updated manually . The salient features of the TOC as a knowledge structure for organizing the knowledge base are summarized in Table 10.

Information Retrieval Mechanisms

To illustrate the use of the conceptual network in retrieving information consider the following example. The stem concept 'heart' may have as its branches the terms 'endocard...', 'myocard...', 'pericard...', 'valves', etc. Likewise the stem concept 'infection' may branch to the terms 'bacterial', 'viral', etc. An article that is stored under the keywords 'subacute bacterial endocarditis' could be searched for by entering 'heart' 'and' 'infection' as criteria for the search. This article citation would then be retrieved since the conceptual network 'knew' that 'heart' included 'endocardium' and 'infection' included 'bacterial'. Branches are treated as alternatives ('or' keywords) to the stem keyword when performing a search. For example, using the 'heart' relation defined above, a search strategy that included 'heart' would also include 'endocardium' 'or' 'myocardium' 'or' 'pericardium' 'or' 'valves',etc. As explained previously a branch keyword in one relation may be the stem keyword of another relation, thus allowing a large hierarchial network of related concepts and keywords to be linked together. When a high level concept is referenced in the network links to lower level concepts are automatically invoked and used to identify related concepts for a user's consideration in formulating seach strategies.

Both the TOC and conceptual network may be made interactive at retrieval time. The conceptual network is based on the mathematics of Boolean 'or' logic when used to formulate search strategies. The TOC uses both Boolean 'or' and 'and' logic. All Boolean operations are performed in a manner that is transparent to the user and need only concern the expert during the design phase of building the knowledge strucutres. By making these structures interactive a non-expert user is prompted to consider related concepts as extensions of the search criteria as well as considering branches that are a subset of a given concept to narrow the search criteria that might otherwise be missed.

Table 10

Features of the Table of Contents structure used in Knowledge Base Organization

1. Provides a framework for the organization and structuring of the content of the data base.
2. Permits the usage of strings of words for the expression of concepts (e.g., Information Systems).
3. Permits the usage of Boolean AND and OR logic.
4. Requires MANUAL indexing. The expert must review each record one at a time. The time required of the expert is proportional to the size of the data base.
5. The data base and knowledge structures must be updated manually when the framework of the data base is changed.
6. The TOC may also be made interactive at retrieval time to guide a user in browsing and searching for information within this knowledge base's organizational structure

<u>Specific Study</u>

Our study question asks if a conceptual network, designed by an expert, can be used to effectively retrieve the set of all citations relevant to a given topic. We used the measures of recall and precision to assess the quality of citation retrieval. Recall is defined as the proportion of relevant citations retrieved from the set of all potentially relevant citations, the truth set. Precision is the proportion of relevant citations in the actually retrieved set of records. Relevance is determined by the medical expert who designs the search strategy.

Building and managing knowledge bases requires the time and efforts of several groups of people. Clerical tasks are easily delegated to clerical staff and may be performed independent of expert supervision. Knowledge engineering tasks require the direct involvement of the expert. The time required to accomplish the knowledge engineering tasks is another dimension to the assessment of the quality of the citation retrieval method.

Using a prototype data base constructed by entering the citations from 18 volumes of Computers and Biomedical Research we have executed the following study. Ten topic areas were identified as being representative of the contents of the journal (see Table 11). The data base contains 824 citations. These were read and reviewed one at a time in order to classify them within the 10 topical content areas of interest. Using this method of manual review we were able to determine the truth set for evaluating the quality of subsequent search strategies. The truth set was our best determination of the set of all relevant citations in the data base for each of the 10 topics. Ten search strategies were devised with the aim of retrieving these truth sets of citations. To formulate these strategies a conceptual network was developed by linking together concepts and keywords through the use of relations. Keywords that appeared in the titles of the articles were available for use within the relations. Additional keywords were also defined by the expert in higher levels of the conceptual network to facilitate the hierarchical structuring of the network. The intellectual challenge was to link keywords together in such a way so as to fully express all of the subtopic areas related to each general topic area. By entering the general topic keyword, i.e., 'cardiology', all related concepts and keywords would be automatically identified and included as

Table 11

Topics Included in the Conceptual Network Study

1.	Biostatistics and Epidemiology
2.	Databases
3.	Data Acquisition
4.	Decision-making
5.	Educational Systems
6.	Hospital Information Systems
7.	Image Analysis
8.	Cardiology
9.	Neurology
10.	Pulmonary

logical 'or' extensions of the search criteria. A search of the data base was then performed using the conceptual network for each of the 10 topic areas. The retrieved sets of citations were then compared against their respective truth sets. False positives and false negatives were identified. We were thus able to assess, by one measure, the quality of the conceptual network as a mechanism for defining and retrieving information.

A potential threat to the external validity of this research design arises from the fact that Dr. Warner determined the criteria for inclusion in the truth sets of citations as well as the design of the conceptual network for retrieving the experimental sets of citations. While it might be felt that this proceedure would artificially inflate the measures of precision and recall, it should be remembered that this design is typical of the types of knowledge base implementations that ILIAD is intended to facilitate. In other words the expert author who develops the knowledge base will also build the knowledge structures that are to be used in retrieving from it relevant citations to a particular topic, as the expert understands the topic. It is one of the basic premises of the program design that the knowledge built into the knowledge structures would reflect the biases of their expert author. These knowledge structures may then be added to and elaborated upon by individual users to accommodate their understandings of topics being researched. It seems appropriate therefore that both our truth standard and our experimentally determined sets of records should reflect the biases of our knowledge base author.

A certain undesireable artifact was identified and removed from the study to ensure that the statistical analyses were correct. This artifact was introduced into the experiment by the secretarial data entry phase of the data base development. Keywords from the titles of articles that are essential for the proper identification of the article's content had in some instances been omitted. This was likely due to keystroke errors and was thought to be significant enough that corrective controls were taken. These entailed the manual reviewing of each article in the truth standard set to ensure that all appropriate keywords from the title were included in the set of keywords used for the indexing of that citation.

CHAPTER IV

RESULTS

Precision and recall have been previously defined (see Specific Study section). The 'false positive' rate is the proportion of citations retrieved that were not actually 'hits'. The 'false negative' rate is the proportion of relevant citations in the data base that were not retrieved. There is no absolute interpretation of the statistics given by recall and precision. They are percentages and their interpretation is based on the magnitude of the search involved. On the average there were 81 citations in the truth sets of each of our 10 topics. As a rule precision and recall are inversly related to each other. By definition the closer that each is to 100% the higher the inferred quality of the search strategy. The statistics from the 10 searches were summed together and an average was computed for each measure. The results are as follows: Recall = 82 %; Precision = 90 %; False Positive rate = 10 %; and False Negative rate = 18 % (see Table 12). The time required of the expert to design and build the conceptual network of relations was approximately 6 hours.

The standard deviation of the precision of the searches = 7. The standard deviation of the recall of the searches = 9. Furthermore the ranges of the precision and recall were 79 - 99 and 65 - 92 respectively. An interesting analysis shows that there was a slightly positive correlation between the recall and precision obtained in our study (R = .42). This is in contrast to the inverse relationship that has been reported in the literature between recall and precision using other search methodologies.[7]

Table 12

Results from the Conceptual Network Study

		Number of Citations	% Precision	% Recall
1.	Biostatistics and Epidemiology	92	80	77
2.	Databases	85	87	84
3.	Data Acquisition	42	79	86
4.	Decision-making	103	87	65
5.	Educational Systems	12	92	92
6.	Image Analysis	75	89	68
7.	Wave Form Analysis	121	99	89
8.	Cardiology	203	94	87
9.	Neurology	35	98	88
10.	Pulmonary	43	97	86
	Mean:	81	90	82
	Standard Deviation:		7	9
	Range:		79-99	65-92

CHAPTER V

DISCUSSION AND CONCLUSIONS

The process of modeling knowledge and making these knowledge models apparent to others is intended to simplify and expedite the process of information retrieval and transfer. Our knowledge models are based on structures that are intended to model the knowledge structures used within the expert's mind. A chief strength of our approach is that it permits an individual to structure a personal data base according to personal preferences. One is not required to adapt to the structuring mechanisms of commercial data bases, etc. This flexibilty allows different experts to develop knowledge models that reflect their own intuitive structuring schemes. An important aspect of this approach to knowledge base management is that the quality of the knowledge base will reflect the amount of effort expended by the expert to develop it. Based on the time and interest that is invested in the designing of the knowledge structures and the indexing of the knowledge contents the user may be able to reap substantial or poor rewards in seeking information within the knowledge base.

The knowledge structures are capable of expressing the organizational schemes underlying the expert's reasoning processes. The ILIAD program allows experts to define rich conceptual networks wherein concepts are defined and linked together as reflections of their understanding of the associations between concepts. This network may in turn facilitate the retrieval of information from the knowledge base as the nonexpert benefits from the knowledge of the expert. By modeling the interconcept links in the expert's network for the user to employ in retrieving information, the probability is higher of any one concept's evoking other, related concepts. It follows that there exists a correspondingly higher probability of retrieving the set of all relevant citations with denser knowledge structures. Using this approach to information retrieval the user is not constrained by the requirement of specifying the exact keywords that were used by the person responsible for storing the citation record. The risk is minimal of missing desired information since each keyword is expanded to include 'related' keywords through the conceptual network.

An important aspect of the knowledge models developed with the structures provided by ILIAD is the ability to up-date and modify them with time. In a sense the knowledge models may evolve with the maturing intellect of their human counterparts. This evolutionary nature of the knowledge model is an important aspect of the challenge of mimicing or modelling the human intellect. Medical students may therefore begin developing a knowledge model and a knowledge base as soon as they begin medical school and further adapt it to their maturing intellect throughout their medical career.

As has been discussed ILIAD has been developed with a goal of modeling the natural knowledge organizational processes of the human mind. This is in keeping with the special challenge that all developers of bibliographic search systems face. Text portions of documents are written in English or some other redundant and ambiguous natural language and there is little doubt that the physician prefers using the same natural language in expressing search topics or concepts of interest. Furthermore,experience has shown that most physicians are unwilling to learn the intricacies and subtleties of access protocols, command languages, Boolean search strategy formulation, and controlled vocabularies.[7] All these issues are addressed in ILIAD's design.

There are several factors to discuss that are pertinent to the search results obtained from our specific study. Of the factors that influence the false negative rates encountered the most significant one deals with the degree to which journal article titles meaningfully convey information about the content of the article itself, a factor outside the control of our study. This is believed to account for a high percentage of the citations missed by the conceputal network approach,lowering the recall rate by approximately 15%. In other words, based on our study there exists approximately 15% of all journal articles that are relevant to a given topic but would not be found without keyword indexing in addition to the use of title keywords.

Another factor that could influence the recall rate of the conceptual network searches and cause it to be artificially high is referred to as the "cumbersome factor" and has to do with human fatigue. Interpreted, it means that while the expert is perfectly capable of accurately identifying all citations within a subject area by manually reviewing them . However, the task is tedious and cumbersome to perform. Articles that fall into one obvious topical classification are sometimes not doubly classified into additional appropriate topical areas. For instance, articles dealing with heart failure and the functioning of the pulmonary system might be indexed under Cardiology but not under Pulmonary. This factor may have had important effects on the results of this study causing the truth standard to be inflated to an undetermined degree in some areas.

The False positive rate of the experimental searches causes concern if the number of citations retrieved incorrectly by the search employing the conceptual network is high. A high false positive rate causes a substantially increased reading burden for the user requiring the filtering out of the false positive citations . This leads to a very inefficient utilization of time when time is already at a premium.

The experimental design was such that the results can only be interpreted in the light of an "expert user" performing searches on the system. From this study one may not make strong inferences as to the application of these concepts in situations where other users attempt to employ another expert's knowledge models, nor to the quality of retrieval that would be realized by their efforts.

The prototype development stage of the ILIAD program has been characterized by a continual implementation-feedback loop for alpha-testing of the knowledge modeling tools and their subsequent refinement. A team of collaborating physicians has been involved in this testing process. It is hoped that as a result of their efforts a set of useful modeling tools with wide application has emerged. This hypothesis must be further tested. I wish to indicate ,however, that this team of physicians has been very crucial to the evolution of the program's design and is responsible for many significant improvements and enhancements to the original concept.

The knowledge modeling tools developed within the program will accommodate the building of personal medical knowledge bases of bibliographic information and facilitate the sharing of this wealth of expert knowledge. Not only does it facilitate the sharing of knowledge but it also permits the sharing of reprint files and other resources. These resources otherwise remain inaccessible to people other than their owner. The knowledge base itself will probably in time be of greater value than the actual physical resource, e.g., personal library and reprint files.

Conclusions of Specific Study

Based on the data from our experiment it is concluded that the use of a conceptual network to retrieve information related to a particular topic is useful in defining concepts and formulating search strategies. Furthermore, searches performed using these knowledge structures yield very high quality retrieval results as measured by recall and precision. The conceptual network approach also decreased the time required both to build the knowledge structures and perform the on-line searching to the point where it represented a reasonable demand of the expert's and the user's time. One of our initial concerns was to be able to minimize the amount of time required by the expert to perform the knowledge engineering tasks while at the same time optimize the quality of citation retrieval. We conclude that the conceptual network approach accomplishes this objective.

Directions for Future Research

The use of computerized tools in assimulating, organizing, and managing the biomedical literature is intended to serve as an aid in combating clinical entropy. They offer the physician an automated means of using the literature to sharpen clinical skills as opposed to being smothered by reports of innovations in diagnosis, prognosis and therapy. [12] There are those who believe that the increasing amount of scientific and technical research being reported in the literature will create a volume of information so large as to frustrate the very purpose for which it was created.[7] If this prediction is not going to become reality, then a larger percentage of our resources must be expended on developing ways of more effectively structuring and handling the mass of information being generated.

It is conceived that eventually specialized knowledge bases will be networked together so that users across an ILIAD network will have access to a rich source of expert medical knowledge on any given medical subspeciality. Not only will these types of networks benefit local users, but they can be so designed that users from geographically remote areas may reap their benefits as well. In order to easily promote the sharing of knowledge across an ILIAD network, due consideration must be given to the need for compatibility and similarity of knowledge models dealing with the same medical disciplines. Such an effort is being undertaken within the Departments of Internal Medicine and Medical Informatics at the University of Utah's Medical School, as additional ILIAD knowledge bases are being developed.

There are many problems attendant to information retrieval and structuring in science and technology that have not been discussed . For example, what criteria should be used to determine whether a datum or statement is a valid entry to a data base. The NLM tends to rely on the published literature. Other systems, including ILIAD, believe that relying on expert opinion is a step forward in the selection process. Regardless of which is used, there immediately arises a second set of questions: What published literature? Whose expert opinion? These problems become even more complex when information in data bases is updated.

ILIAD has been designed with the anticipation that eventually the criteria for ensuring only valid entries to the data base will be developed. Software features are now built into the program to facilitate the critical appraisal process; however, much work remains to be done in the area of defining these critical appraisal criteria. Additionally, significant thought must be given to the process of training and educating potential experts and users in critical appraisal skills. The pursuit of these issues should nicely complement the screening process that already adds significant value to an ILIAD knowledge base as a repository of knowledge. These are issues at the forefront of medical informatics' applications in automated information retrieval and define areas of further research in the development of effective systems.

Finally, the ILIAD project is now at a point of possible departure where interesting studies may be devised to test its effectiveness among a wide variety of health care provider "users." As a prototype of medical informatics' applications to education, the need of and potential impact of this type of technology integrated into the medical curriculum should be carefully studied and assessed. Can this type of information management tool positively impact the teaching of medicine?

Summary

Well-validated knowledge, available in the clinical literature, is not known or not used by all practicing physicians. We assume that this information, if applied to patient care, may make a difference in the outcome of a patient's illnesses and the effective and efficient use of health care resources. ILIAD as a tool for developing screened, personal knowledge bases may become part of the practice of medicine. The ILIAD program offers a means of effectively modeling and disseminating the knowledge of medical experts. It offers a new means of transfering knowledge to medical students as the medical expert serves as mentor and teacher. This type of computer application to knowledge management, designed to match the cognitive processes of an expert on a given subject, may offer the physician an effective means of remaining closer to the frontier of enlightened medical practice.

APPENDIX A

PROGRAM MENUS

MAIN MENU	UTILITIES MENU
A)...A)DD CITATIONS B)...ADD B)OOK Citations S)...S)EARCHER X)...eX)it Program	C)...C)reator D)...D)eleter E)...E)ditor P)...P)rinter U)...U)pDate

SELECTION —>__

Figure 5: Program's Main Menu

```
|--------SEARCHER---------- |----------BROWSER----------- |-----------ANALYZER---------- |
| 1. SEARCH by TTOC         | 3. BROWSE TTOC              | 6. ANALYZER                  |
| 2. SEARCH by INDEX        | 4.  BROWSE INDEX            |                              |
|                           | 5. BROWSE MODIFIERS         |                              |
| R)...Return to Main Menu  |                             |                              |
|---------------------------|-----------------------------|------------------------------|

                    CHOOSE OPTION -> __
```

Figure 6: SEARCHER Menu

```
--------------CREATOR--------------
 1.  TTOC MAIN HEADING
 2.  TTOC SUBHEADINGS
 3.  MAIN MODIFIER
 4.  SUBMODIFIERS
 5.  RELATION
 6.  ABBREVIATION

 R)..R)ETURN to MAIN MENU
--------------------------------------------

CHOOSE CREATE OPTION -> __
```

```
-----------------DELETER---------------
 1.  RELATION

 R)..R)ETURN to MAIN MENU
--------------------------------------------

CHOOSE OPTION -> __
```

Figure 7: CREATE and DELETE Menus

```
---------------EDITOR----------------
 1.  CITATION
 2.  TTOC
 3.  RELATION
 4.  MODIFIERS
 5.  ABBREVIATIONS TABLE
 6.  DATA DICTIONARY

 R)..R)ETURN to MAIN MENU
-------------------------------------

CHOOSE EDIT OPTION -> __
```

FIGURE 8: EDIT Menu

```
---------------PRINTER----------------
 1.  CITATIONS
 2.  TTOC
 3.  RELATIONS
 4.  MODIFIERS
 5.  ABBREVIATIONS TABLE
 6.  INDEX
 7.  FILES: Level A

 R).. R)ETURN TO MAIN MENU
----------------------------------------

    CHOOSE PRINT OPTION -> __
```

```
-------------DESTINATION-------------
          1. TERMINAL
          2. PRINTER
---------------------------------------------

    CHOOSE OPTION  -> __
```

Figure 9:   PRINTER Menus

APPENDIX B

TTOC OF THE MEDICAL INFORMATICS' KNOWLEDGE BASE

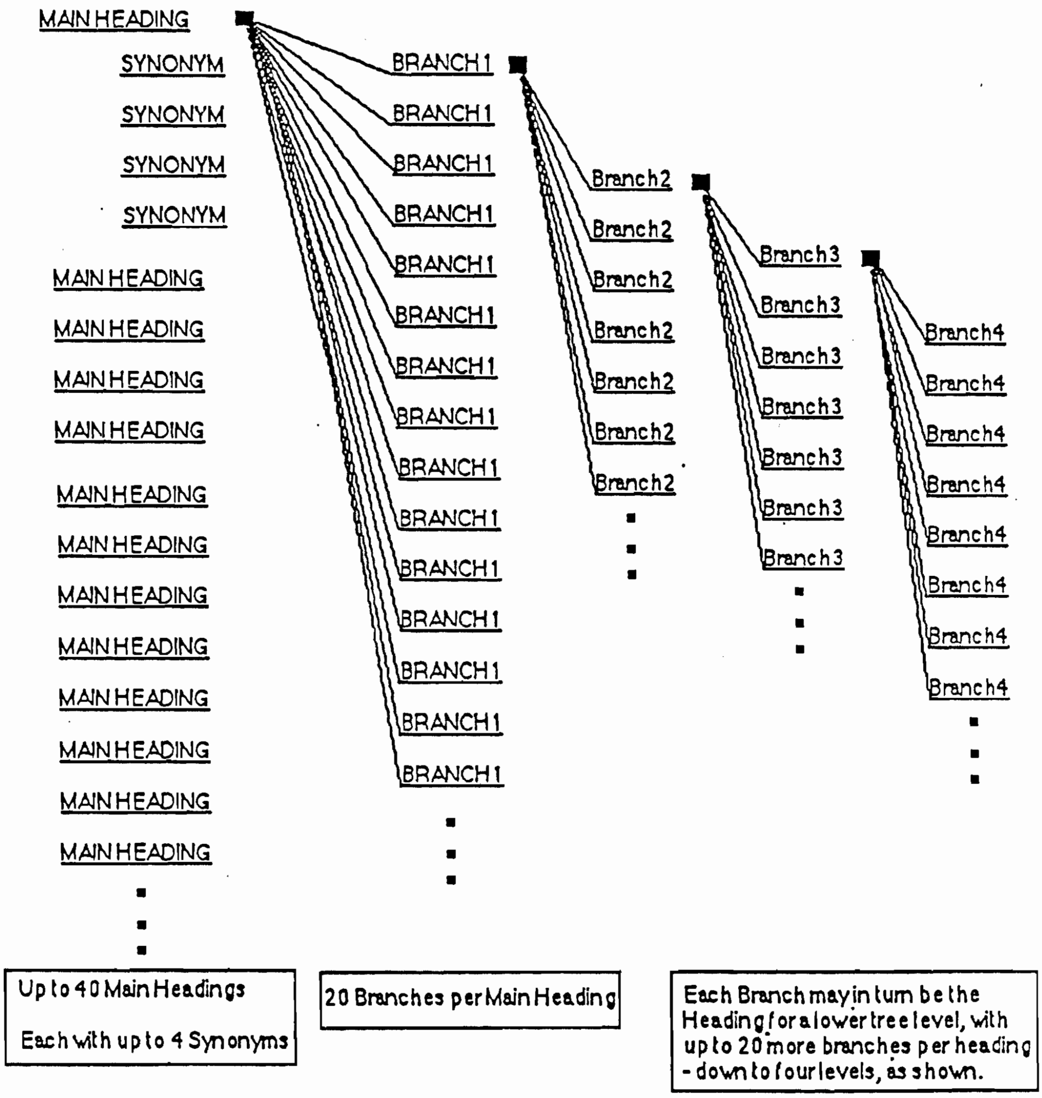

Figure 10: Schematic of the TTOC Structure

TOPICAL TABLE OF CONTENTS

FILE: MEDICAL INFORMATICS PAGE 1 OF 1

*1. DST	86 :	DECISION SUPPORT TECHNOLOGY
2. SIMULATION/MODELING	218:	
*3. WAVE FORM PROCESSING	121:	
*4. PICTURE PROCESSING	59 :	
*5. DATA BASE	66 :	
6. BIOSTAT/EPIDEMIOLOGY	83 :	
7. RESEARCH	1 :	
8. INSTRUMENTATION	2 :	
9. DATA ACQUISITION	32 :	
10. DATA DISPLAY/REPORT	7 :	
11. LANGUAGE PROCESSING	6 :	
*12. EDUCATION SYSTEMS	4 :	
*13. COMPUTER SYSTEMS	27 :	
*14. AREA OF MEDICINE	154:	

CHOOSE TOPIC HEADING #'S : [Separated be BLANKS] <NONE> ->__

Figure 11: TTOC Menu for Prompting

1. DST : DECISION SUPPORT TECHNOLOGY
    - DECISION TYPE
        - DX : DIAGNOSIS
        - PX : PROGNOSIS
        - RX : TREATMENT
    - FUNCTIONAL FEATURES
        - DATABASE
        - EXPLANATION
        - DATA SOURCE
            - TERMINAL
            - INSTRUMENTATION
        - INITIATES DECISIONS
            - RECIPIENT
            - DATA DRIVEN
        - EASY TO MAINTAIN
    - OPERATIONAL STATUS
        - CLINICALLY TESTED
        - STILL IN USE
        - COMMERCIAL PRODUCT
    - TECHNOLOGICAL DATA
        - KNOWLEDGE REPRESENTATION
            - IF-THEN
            - PROBABILITY
            - DISCRIMINANT FUNCTIONS
            - DECISION TREE
            - OTHER
        - IMPLEMENTATION
            - HARD CODE
            - FRAMES
            - LISTS

2. SIMULATION/MODELING

3. WAVE FORM PROCESSING
    - ECG
        - RHYTHM
            - NODAL TACHYCARDIA
            - ATRIAL FIBRILLATION
            - ATRIAL FLUTTER
            - VP CONTRACTION
            - VENTRICULAR TACHYCARDIA
            - 2ND DEGREE
            - 3RD DEGREE
        - MORPHOLOGY
            - MI : MYOCARDIAL INFARCTION
            - HYPERTROHY
    - PRESSURE WAVE FORMS
    - INDICATOR DILUTION CURVES
    - EEG

Figure 12: Medical Informatics' Knowledge Base's TTOC

4. PICTURE PROCESSING
   - DIAGNOSTIC IMAGING
   - STORAGE SYSTEMS

5. DATABASE
   - MIS : MEDICAL INFORMATION SYSTEMS
     - ADMINISTRATION
     - FINANCIAL
     - ADT
     - MEDICAL RECORDS
       - DRG
       - ICD-9
     - NURSING APPLICATIONS
     - RADIOLOGY
     - LABORATORY
     - PHARMACY
     - ICU
     - HEART LAB
   - DESIGN
   - MANAGEMENT
   - MODEL
     - HIERARCHICAL
     - RELATIONAL

6. BIOSTAT/EPIDEMIOLOGY

7. RESEARCH

8. INSTRUMENTATION

9. DATA ACQUISITION

10. DATA DISPLAY/REPORT

11. LANGUAGE PROCESSING : NATURAL LANGUAGE PROCESSING

12. EDUCATION SYSTEMS
    - CONTINUING MEDICAL EDUCATION
    - CLINICAL EDUCATION
    - RADIOLOGY EDUC. SYSTEMS

Figure 12 (continued)

13. COMPUTER SYSTEMS
    - HARDWARE
        - MICRO
        - MINI
        - MAINFRAME
    - SOFTWARE
        - LANGUAGES
        - OPERATING SYSTEMS
    - COMMUNICATIONS
        - LAN : LOCAL AREA NETWORKS
        - DISTRIBUTED NETWORKS
        - INTERFACING MACHINES
        - TELEMATICS

14. AREA OF MEDICINE
    - MEDICINE
        - GENERAL
        - CARDIOLOGY
        - DERMATOLOGY
        - ENDOCRINOLOGY
        - GASTROENTEROLOGY
        - NEUROLOGY
        - PULMONARY
        - RHEUMATOLOGY
        - HEMATOLOGY
    - SURGERY
        - GENERAL
        - ABDOMINAL
        - CARDIOVASCULAR
        - OPHTHALMOLOGY
        - EAR-NOSE-THROAT
        - UROLOGY
    - OB/GYN
    - PEDIATRICS
    - PSYCHIATRY

Figure 12 (continued)

APPENDIX C

RELATIONS OF THE MEDICAL INFORMATICS' KNOWLEDGE BASE

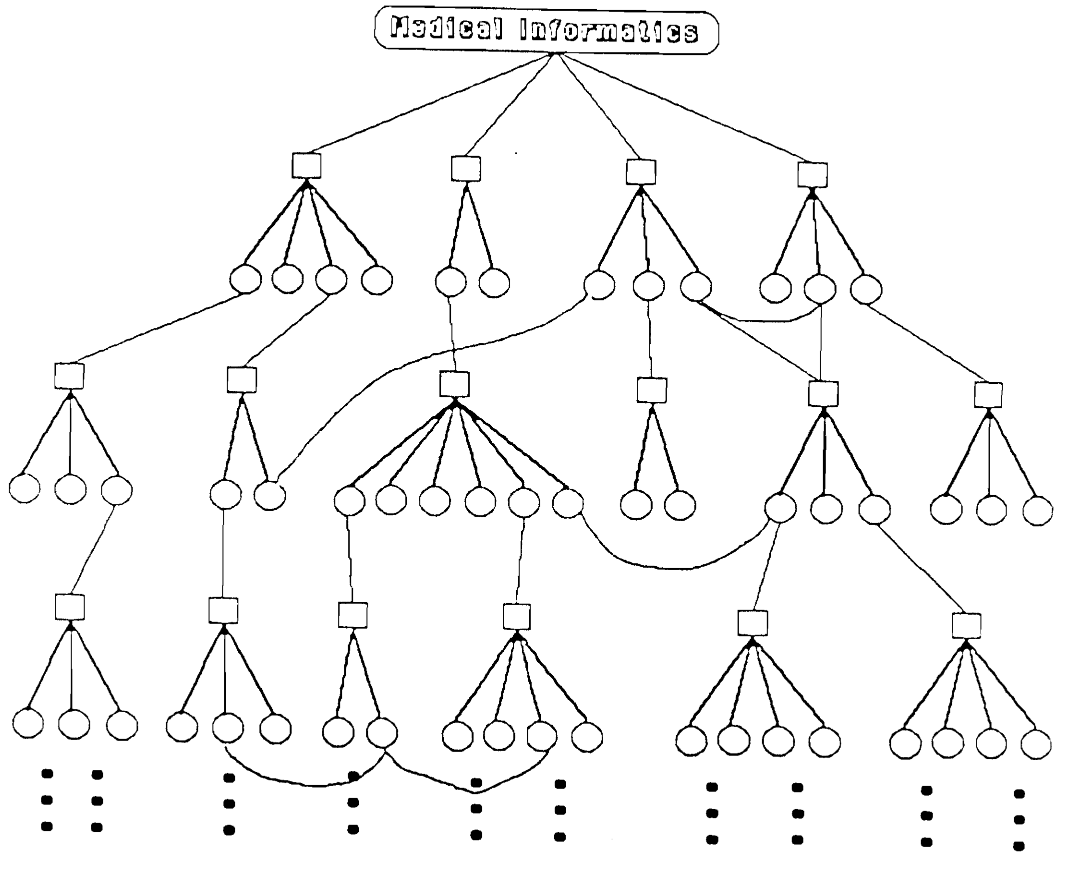

Figure 13: Schematic of Conceptual Network Tree Structure

TRUNK:	RHYTHM	
SYNONYMS:		
BRANCHES:		ATRIAL VENTRICULAR FIBRILLATION TACHYCARDIA FLUTTER NORMAL
TRUNK:	WAVE	
SYNONYMS:		
BRANCHES:		PRESSURE ECG EEG INDICATOR DILUTION FILTER FOURIER TRANSFORMS SPIRO ELECTROEN
TRUNK:	ECG	
SYNONYMS:	EKG ELECTROC	
BRANCHES:		ELECTRODE
TRUNK:	STATISTICS	
SYNONYMS:		
BRANCHES:		BIOSTAT EPIDEMIOLOGY

Figure 14: RELATIONS from the Medical Informatics' Knowledge Base

TRUNK: BIOSTAT

SYNONYMS:

BRANCHES:

PROBAB
LEAST-SQUARES
GOODNESS-OF-FIT
REGRESSION
STEPWISE
RANDOM
DATA -ANALYSIS
BAYES
PIECEWISE
LINEAR
NONLINEAR
MATRIX
INTERPOLATION
STOCHASTIC
DETERMINISTIC
QUANTITATIVE
CORRELATION
GOODNESS-OF-FIT
SENSITIVITY
FOURIER
PRECISION
RATIO

TRUNK: EPIDEMIOLOGY

SYNONYMS:

BRANCHES:

CLINICAL TRIALS
PREVALENCE
INCIDENCE
LIKELIHOOD
RISK-FACTORS
MORTALITY

Figure 14 (continued)

TRUNK:	ARTERY	
SYNONYMS:	ARTERIES ARTERIAL	
BRANCHES:		AORTIC AORTA RADIAL BRACHIAL FEMORAL PULMARY HEPATIC CORONARY CAROTIC
TRUNK:	INDICATOR	
SYNONYMS:	TRACER MARKER	
BRANCHES:		ISOTOPE DYE RADIOACTIVE
TRUNK:	CAI	
SYNONYMS:		
BRANCHES:		EDUCATION INSTRUCTION LEARNING TRAINING STUDENT TEACH
TRUNK:	DATA BASE	
SYNONYMS:	DATA-BASE DATABANK	
BRANCHES:		MIS DATA BASE DESIGN DATA BASE MODELS DATA BASE MODELS MANAGEMENT

Figure 14 (continued)

TRUNK:	MIS	
SYNONYMS:		
BRANCHES:		RECORD SYSTEM

TRUNK:	DATA BASE MODELS	
SYNONYMS:		
BRANCHES:		RELATIONAL HIERARCHICAL

TRUNK:	MANAGEMENT	
SYNONYMS:		
BRANCHES:		STORAGE RETRIEVAL REPRESENTATION REPORT

TRUNK:	VESSEL	
SYNONYMS:		VASCULATURE
BRANCHES:		VEIN VENOUS ARTER

Figure 14 (continued)

TRUNK: DATA ACQUISITION

SYNONYMS:

BRANCHES:

DATA COLLECTION
DATA ENTRY
DATA MANAGEMENT
ACQUISITION
REPORTING
DISPLAY
ENTRY
MARK SENSE
ON-LINE
MONITORING
INTERVIEW
HISTORY
QUESTIONNAIRE

TRUNK: IMAGING

SYNONYMS:

BRANCHES:

HIGH RESOLUTION
HIGH-DENSITY
HIGH-FREQUENCY
PATTERN RECOGNITION
PICTURE ANALYSIS
PIXELS
RECONSTRUCTION
X-RAY
RADIOLOGY
CAT SCAN
TOMOGRAPHY
ULTRASOUND
NMR
MRI
NUCLEAR MAGNETIC
GRAPHIC
VIDEO
DRAWING
TELEVISION
CONTOURS
XEROXMAN

Figure 14 (continued)

TRUNK: DECISION

SYNONYMS:

BRANCHES:
CLASSIFICATION
DIAGNOS
DETERMIN
IDENTIFICATION
INTERPRETATION
PROGNOS
ADVISE
DETECT
ALARM
ASSESS
PREDICT
CONSULT

TRUNK: CARDIOLOGY

SYNONYMS:

BRANCHES:
HEART
VESSEL
ECG
ANGIO
ARRHYTHMIA
BLOCKS

Figure 14 (continued)

TRUNK:	NEUROLOGY	
SYNONYMS:		
BRANCHES:		NEURAL NETS NERVE SLEEP EVOKED CEREBRAL EMG REFLEX BRAIN

TRUNK:	PULMONARY	
SYNONYMS:		
BRANCHES:		LUNG EXPIRATORY SPIRO RESPIRATION BLOOD-GAS OXYHEMOGLOBIN BREATH VENTIL CARDIORESPIRATORY ALVEOLAR

Figure 14 (continued)

SELECTED BIBLIOGRAPHY

1 Dornhorst AC: Information overload: Why medical education needs a shakeup. *Lancet* 1981; 2: 513-514.

2 Anderson J, Graham A: A problem in medical education: Is there an information overload? *Medical Education* 1980; 14: 4-7.

3 Lindberg DAB: *The growth of medical information systems in the United States.* Lexington, Massachusetts: Lexington Books, 1979.

4 Ball MJ: Medical information systems in the USA. Lecture notes in *Medical Informatics* 1981; 11: 22-32.

5 Sacket DL, Haynes RB, Tugwell P: *Clinical epidemiology: A basic science for clinical medicine.* Boston, Massachusetts: Little, Brown and Company, 1985.

6 Levinson D: Information, computers, and clinical practice. *JAMA* 1983;249: 607-609.

7 Doszkocs TE, Rapp BA, Schoolman HM: Automated information retrieval in science and technology. *Science* 1980; 200: 25-30.

8 Nau DS: Expert computer systems.*Computer* 1983; 63-85.

9 Schoolman HM, Bernstein LM: Computer use in diagnosis, prognosis, and therapy: Computers are used to simulate the reasoning of expert clinicians.*Science* 1978;200: 926-931.

10 Reichertz PL: Medical informatics - fiction or reality? *Methods Inf. Med.* 1980; 19: 11-15.

11 Muller S (Chairman): *Physicians for the twenty-first century: The GPEP report.* Washington, D.C.: Association of American Medical Colleges, 1984.

12 Sackett DL, et al.: How to read clinical journals: I. Why to read them and how to start reading them critically. *CMA Journal* 1981; 124: 555-558.

13 Stinson ER, Mueller DA: Survey of health professionals' information habits and needs, conducted through personal interviews. *JAMA* 1980; 243: 140-143.

14 Adams S: The MEDLARS system. Presented at the 6th Annual meeting of the Conference of Biological Editors, Washington, D.C., March 20, 1963